WHAT IS CLAIMED IS:

1. A method for analyzing nucleotide sequence information during haplotyping analysis, the method comprising:

selecting a data superset comprising single nucleotide polymorphism (SNP) information describing a plurality of SNPs, each SNP associated with a plurality of haplotypes, wherein each haplotype is determined by the sequence of SNPs present in the haplotype;

identifying groupings of analogous SNPs from the data superset whose sequences are analogous in two or more haplotypes;

selecting at least one representative SNP from each grouping of analogous SNPs to be included in a reduced data subset; and

performing a haplotyping analysis using the reduced data subset.

2. The method of Claim 1 further comprising, selecting at least one non-analogous SNP to be included in the reduced data subset.

3. The method of Claim 1 wherein, performing the haplotyping analysis using the reduced data subset substantially preserves haplotype diversity.

4. The method of Claim 1 wherein, performing the haplotyping analysis using the reduced data subset requires fewer computations to complete relative to performing haplotyping analysis using the data superset.

5. The method of Claim 4 wherein, computational performance during haplotyping analysis is improved using the reduced data subset.

6. The method of Claim 1 wherein, the haplotyping analysis comprises discriminating between haplotypes associated with the SNP information.

7. The method of Claim 1 further comprising,

identifying at least one diversity subset from the reduced data subset comprising a plurality of SNPs associated with a selected haplotype;

identifying combinations of SNPs selected from the at least one diversity subset and calculating an entropy value for each SNP combination;

identifying a refined diversity subset from the SNP combinations having an entropy value within a selected range and a selected number of SNPs; and

performing the haplotyping analysis using the refined diversity subset.

8. The method of Claim 7 wherein, the entropy value for each diversity subset is determined by assessing the relative frequency of occurrence of the selected haplotype.

9. The method of Claim 7 wherein, the entropy value for each diversity subset is determined using a Shannon entropy determination.

10. A method for analyzing nucleotide sequence information, the method comprising:

selecting a data superset comprising single nucleotide polymorphism (SNP) information describing a plurality of SNPs, each SNP associated with a plurality of haplotypes, wherein each haplotype is determined by the sequence of SNPs present in the haplotype;

identifying regions of analogous SNP information for each of the plurality of haplotypes;

identifying at least one representative SNP from the analogous SNP information for each region;

forming a reduced data subset wherein at least a portion of the analogous SNP information is excluded from the reduced data subset while haplotype diversity is preserved by inclusion of the at least one representative SNP in the reduced data subset; and

performing the haplotyping analysis using the reduced data subset.

11. The method of Claim 10 wherein, identifying analogous SNPs comprises identifying groups of two of more SNPs whose sequences are substantially identical for each of the plurality of haplotypes.

12. The method of Claim 10 wherein, identifying analogous SNPs comprises identifying groups of two of more SNPs whose sequences are substantially complimentary for each of the plurality of haplotypes.

13. The method of Claim 10 wherein, selecting at least one representative SNP comprises excluding substantially all of the analogous SNP information for each region with the exception of the at least one representative SNP identified from each region.

14. The method of Claim 10 wherein, analogous SNPs are identified by comparing the SNP information in a pairwise manner to identify SNPs whose sequence is identical or complimentary in two or more haplotypes.

15. The method of Claim 10, wherein performing the haplotyping analysis using the reduced data subset provides similar haplotyping diversity information as the data superset from which it was derived while improving computational performance during haplotyping analysis.

16. The method of Claim 10, wherein use of the reduced data subset during haplotyping analysis reduces the computational complexity of performing the haplotyping analysis.

17. The method of Claim 16, wherein formation of the reduced data subset reduces the computational complexity of performing of haplotyping analysis by reducing the total number of SNPs to be analyzed.

18. The method of Claim 10, wherein the reduced data subset is used in the evaluation of a selected genetic loci.

19. The method of Claim 10, wherein haplotyping analysis using the reduced data subset facilitates discrimination between haplotypes with substantially the same degree of specificity as the data superset from which they were derived.

20. The method of Claim 10, further comprising:

identifying a plurality of diversity subsets, each comprising one or more SNPs associated with a selected haplotype, by selecting combinations of SNPs associated with the selected haplotype;

calculating an entropy value for each diversity subset and comparing these values to the entropy value determined for the diversity subset containing all associated SNPs;

identifying an refined diversity subset from the reduced data subset having substantially the greatest entropy value and least number of associated SNPs; and

performing the haplotyping analysis using the refined diversity subset.

21. The method of Claim 20, wherein the entropy value for each diversity subset is determined as a probability factor defined for each associated haplotype wherein the probability factor describes the relative associated frequency of occurrence of the selected haplotype.

22. The method of Claim 21, wherein the entropy value is calculated using a Shannon entropy determination.

23. The method of Claim 24, wherein the diversity subset having the greatest number of SNP combinations is used as a threshold for determination of the refined diversity subset.

24. The method of Claim 25, wherein the threshold reduces the complexity of calculations in determining the refined diversity subset.

25. A system for analyzing nucleotide sequence information during haplotyping analysis, the system comprising:

a data collection component that provides functionality for selecting a data superset comprising single nucleotide polymorphism (SNP) information describing a plurality of SNPs, each SNP associated with a plurality of haplotypes, wherein each haplotype is determined by the sequence of SNPs present in the haplotype;

a first data analysis component that provides functionality for identifying a plurality of diversity subsets, each comprising one or more SNPs associated with a selected haplotype, by selecting combinations of SNPs associated with the selected haplotype;

a first computational component that provides functionality for calculating an entropy value for each diversity subset and comparing the

resulting entropy values to an entropy value determined for the diversity subset containing substantially all associated SNPs;

a second data analysis component that provides functionality for identifying an refined diversity subset from the data superset having substantially the greatest entropy value and least number of associated SNPs; and

a second computational component that provides functionality for performing the haplotyping analysis using the refined diversity subset.

26. The system of Claim 25, wherein the first computational component determines the entropy value for each diversity subset using a probability factor defined for each associated haplotype wherein the probability factor describes the relative associated frequency of occurrence of the selected haplotype.

27. The system of Claim 25, wherein the first computational component calculates the entropy value using a Shannon entropy determination.

28. The system of Claim 27, wherein the second data analysis component identifies the diversity subset having the greatest number of SNP combinations which is used as a threshold for determination of the refined diversity subset.

29. The system of Claim 28, wherein use of the threshold reduces the complexity of calculations in determining the refined diversity subset.

30. A method for analyzing nucleotide sequence information during haplotyping analysis, the method comprising:

selecting a data superset comprising single nucleotide polymorphism (SNP) information describing a plurality of SNPs, each SNP associated with a plurality of haplotypes, wherein each haplotype is determined by the sequence of SNPs present in the haplotype;

performing a first data reduction on the data superset by identifying redundant SNPs comprising two of more SNPs whose sequences are identical or complimentary for each of the plurality of haplotypes and removing at least a portion of the redundant SNPs from the data superset;

performing a second data reduction on the data superset by comparing the SNP information in a pairwise manner to identify analogous SNPs whose sequences are identical in two or more haplotypes and removing at least a portion of the analogous SNPs; and

performing a haplotyping analysis using the remaining SNP information in the data superset.

31. The method of Claim 30, further comprising performing a third data reduction using the remaining SNP information wherein the third data reduction comprises:

identifying a plurality of diversity subsets, each comprising at least one SNP associated with a selected haplotype, by selecting combinations of SNPs associated with the selected haplotype;

calculating entropy values for each diversity subset;

comparing the calculated entropy values to an entropy value determined for a diversity subset containing substantially all associated SNPs;

identifying a refined diversity subset having substantially the greatest entropy value and least number of associated SNPs; and

performing the haplotyping analysis using the refined diversity subset.

32. The method of Claim 31, wherein the entropy value for each diversity subset is determined based upon a probability factor defined for each associated haplotype wherein the probability factor describes the relative associated frequency of occurrence of the selected haplotype.

33. The method of Claim 31, wherein the entropy value is calculated using a Shannon entropy determination.